Biol Theory (2015) 10:134–144
DOI 10.1007/s13752-015-0206-2

ORIGINAL ARTICLE

On the Free-Rider Identification Problem

Ronald J. Planer[1]

Received: 26 July 2014 / Accepted: 15 March 2015 / Published online: 8 April 2015

Abstract Samuel Bowles and Herbert Gintis have argued that individual-selection accounts of human cooperation flounder in the face of the free-rider identification problem. Kim Sterelny has responded to this line of argument for group selection, arguing that the free-rider identification problem in fact poses no theoretical difficulty for individual-selection accounts. In this article, I set out to clarify Bowles and Gintis' argument. As I see matters, the real crux of their argument is this: solving the free-rider identification problem, even in modestly sized social groups, requires that group members are disposed to share social information with one another. The difficulty for individual-selection accounts, according to Bowles and Gintis, is that these accounts have no explanation for why individuals should be disposed to behave in this way. Having clarified their argument, I then turn to Sterelny's criticism, and argue that Sterelny underestimates the challenge being raised by Bowles and Gintis. More specifically, I argue that it is unclear whether the expected benefits of having a disposition to share social information would have outweighed the expected costs for an individual belonging to a Pleistocene social group. Importantly, this is not to say that I am persuaded by Bowles and Gintis' argument; on the contrary, what I claim is that more theoretical (and in particular) empirical work is necessary before the issues under discussion can be settled. I formulate some specific questions which I think future research in this area should aim to address.

Keywords Communication networks · Group selection · Human evolution · Information sharing · Indirect reciprocity · Partner choice · Reciprocal altruism · Social information

Introduction

In their book *A Cooperative Species* (2011), Samuel Bowles and Herbert Gintis launch a sustained attack against individual-selection accounts of human cooperation. Their attack is multifaceted and is backed by an impressive range of modeling, experimental, and archeological evidence. Here my aim is to carefully consider but a single strand in their overall argument.

Bowles and Gintis argue that individual-selection accounts of human cooperation flounder in the face of the free-rider identification problem. This is the problem of identifying individuals who have defected (so as to secure personal gains) in a cooperative interaction, and making this information widely known throughout the group. In order for cooperation to arise and remain stable in even modestly sized, non-kin-based groups, it is imperative that free riders (and more or less only free riders) be punished. However, this cannot happen unless free riders are reliably picked out and collectively designated as such by the group.

Kim Sterelny (2012) has recently responded to this line of argument for group selection, arguing that the free-rider identification problem is a "pseudo-problem" (p. 181). What exactly does Sterelny mean by this? We can be sure he isn't claiming that the free-rider identification problem, at least as I have just described it, is not one that actually exists. Instead, what I take Sterelny to be claiming is just that, *pace* Bowles and Gintis, the free-rider identification

✉ Ronald J. Planer
planer.ronald.jon@gmail.com

[1] Department of Philosophy, Center for Cognitive Science, Rutgers University, New Brunswick, NJ, USA

problem poses no real theoretical difficulty for individual-selection accounts of human cooperation. There is no reason to think that groups in which this problem is solved are necessarily ones in which individuals have had their psychologies shaped by a process of group selection.

I find Bowles and Gintis' argument from the free-rider identification problem to group selection interesting but opaque. In this article, my first task is to clarify their argument. I suggest that the real crux of their argument is this: even in modestly sized social groups, the free-rider identification problem can only be solved provided that group members are disposed to share social information. The difficulty individual-selection accounts confront, according to Bowles and Gintis, is that these accounts have no explanation for why individuals should be disposed to behave in this way. Having clarified their argument, I then consider its plausibility in light of Sterelny's criticism. While there is much in Sterelny with which I agree, I think he ultimately underestimates the challenge being raised by Bowles and Gintis for individual-selection accounts. More specifically, I argue that it is unclear whether the expected benefits of having a disposition to share social information would have outweighed the expected costs for an individual belonging to a Pleistocene social group. None of this is to say that I am ultimately convinced by Bowles and Gintis' argument. Rather, my view is that more theoretical (and in particular) empirical work must be carried out before this debate can be settled, and I conclude by formulating some specific questions that I think future research in this area should aim to address.

Bowles and Gintis on the Free-Rider Identification Problem

It is widely held that kin selection is a non-starter when it comes to explaining human cooperation; the reason being, the groups our Pleistocene ancestors lived in are believed to have contained too many non-kin for this selection mechanism to have been effective. Hence, in the view of most researchers, we are left with two main individual-selection mechanisms for explaining cooperation, namely, reciprocal altruism (Trivers 1971; Axelrod and Hamilton 1981) and indirect reciprocity (Sugden 1986; Alexander 1987; Nowak and Sigmund 1998), both construed broadly.[1]

Bowles and Gintis claim that, for even modestly-sized social groups, explanations based upon either individual-selection mechanism (including some combination of the two) break down.

The following passage offers a nice summary of Bowles and Gintis' view on reciprocal altruism. They write:

> The reason for the ineffectiveness of reciprocal altruism for groups with several members is simple. In groups of two, a free-rider cannot go undetected because a player's payoff reveals the other player's behavior. Equally important, when one member defects in order to punish a Defector, the punishment is uniquely targeted on the Defector. But, in groups larger than two, a player cannot infer who has defected from the knowledge of his own payoff. Moreover, a retaliatory defection punishes not only the initial defector, but also all other members of the group. Moreover, other group members may not have observed the initial defection and hence may think that a retaliatory defection is a free-riding defection, inviting further retaliatory defections. (p. 64)[2]

There are two distinct issues being raised by Bowles and Gintis here. The first concerns the *identification* of free riders; the second, the *targeting* of free riders (and only free riders) for punishment. Although the second issue is also of critical importance to the more general debate between individual-selection and group-selection proponents in this area, I will set it aside in what follows.

Bowles and Gintis point out that when a cooperative interaction (with "simultaneous" moves) involves more than two interactants, there is no guarantee that a given interactant will know who has cooperated and who has defected, as this information cannot be ascertained from one's payoff. In this regard, such interactions are importantly different from dyadic ones. Bowles and Gintis also point out that uncertainty about who has cooperated and who has defected in the context of past interactions can have a kind of ripple effect. For example, if an onlooker A is uncertain as to whether B defected against C in the past, then there is no guarantee that A will be able to tell whether C's current defection against B constitutes a form of free riding or a form of punishment. Bowles and Gintis take these considerations to cast doubt on the claim that reciprocal altruism can offer an adequate explanation for human cooperation. But why, exactly?

Before delving into this question, let us have a look at what Bowles and Gintis say regarding indirect reciprocity. They write:

[1] By "construed broadly" here, I simply mean explanations revolving around repeated interaction (in the case of reciprocal altruism) and ones revolving around reputation (in the case of indirect reciprocity). Below, we will look at a third type of individual-selection mechanism, namely, partner choice, which also relies crucially upon reputation. In this regard, partner choice is closely tied to indirect reciprocity, and I wait until later on in this article to fully distinguish the two.

[2] All page numbers following block quotes in this section refer to Bowles and Gintis (2011).

> [Indirect reciprocity models have] very demanding informational requirements. [Their] ability to target defectors makes [them] an improvement on the repeated game models … but this comes at a steep price. Each individual must know the current standing of each member of the group, the identity of each member's current partner, and whether each individual cooperated or defected against his current partner, since this information is necessary to ascertain the status of one's partner in the indirect reciprocity game. But real-world dyadic interactions are often private, and hence are unlikely to be directly observed by more than a small number of others, vitiating the model for groups of any significant size. (p. 70)

In this passage, we see Bowles and Gintis' objecting to the (in their view) demanding "informational requirements" of indirect reciprocity. Surely, it is reasonable to be more impressed by some of these requirements than others. For example, while knowing the identity of one's current partner(s) is probably trivial in the vast majority of real-life cases, knowing the status of one's partner(s) (whether they are in "good" or "bad" standing in the group), or the status of the individuals one is watching interact, is definitely not trivial. Bowles and Gintis are bothered by something else: they think that, even in small-scale societies, a significant number of cooperative interactions take place in the presence of few or no onlookers.[3] But again, why exactly do Bowles and Gintis take these considerations to undermine explanations of human cooperation based upon indirect reciprocity?

To be sure, there is an uncharitable way of interpreting both sets of remarks here. It can seem like what Bowles and Gintis are claiming is that the free-rider identification problem is effectively insoluble in even modestly sized social groups. Or put more fully: repeated interaction ceases to support cooperation in such groups as interactants will often be left not knowing who has behaved cooperatively and who has behaved selfishly. In addition, reputational considerations cease to support cooperation because there is just far too much for group members to keep track of. The representations group members possess of the past behavior of others are bound to be significantly incomplete at best, implying that free riders will often be treated as cooperators (and perhaps cooperators as free riders). Consequently, humans' disposition to cooperate cannot be explained by appeal to the individual benefits of such a disposition in the evolutionary past. An individual belonging to a Pleistocene social group would have been *worse off*, not *better off*, as a result of having a disposition to cooperate, as free riding would have often gone undetected, and so unpunished. Hence, to explain why humans nevertheless evolved such a disposition, we must appeal to selection operating at the level of groups.

The reason I say this would be an uncharitable interpretation of their argument is simple (although I do not think Bowles and Gintis do enough to discourage this reading of their argument). It is not just a problem for individual-selection accounts of cooperation if free riders cannot be reliably detected. That is, it is also crucially important, if not equally so, for a group-selection account such as Bowles and Gintis' that the free-rider identification problem can be solved in groups like the ones our Pleistocene ancestors lived in. A free rider in a group of cooperators will have a higher fitness, and so free riding will tend to spread through the group unless free riding is identified and punished. So, even if a group consisted solely of cooperators at some initial point in time (in which case, there wouldn't actually *be* any free riders to identify at that time) an inability to solve the free-rider identification problem would inevitably give way to what Richard Dawkins has colorfully described as "subversion from within" (Dawkins 1976). That is, "mutation" would eventually serve to introduce free riders into the group, and once introduced, such individuals would begin to overtake the group. Consequently, levels of cooperation would begin to decay and eventually collapse altogether. Bowles and Gintis are surely aware of this kind of argument against group-selection accounts of cooperation, as it has long been held up as group selection's Achilles' heel. In a word: group-selection proponents such as Bowles and Gintis need a viable explanation for how the free-rider identification problem gets solved every bit as much as their individual-selection opponents.

Fortunately, I think there is a much more plausible way to interpret Bowles and Gintis' argument, one that is suggested by what they have to say on the topic of *informational sharing* between group members. Still discussing the (purported) shortcomings of explanations of cooperation based upon indirect reciprocity, they write:

> … if individuals are entirely self-regarding, they have no reason to report truthfully [the behavior] they have observed. Though an active area of research, explanations of how private information could be converted to accurate public information in a population of amoral self-regarding individuals have not been presented.

[3] Ideally, one would like to see claims about the frequency of (more or less) private cooperative interactions buttressed by concrete ethnographic evidence. Unfortunately, there is not space to consider the ethnographic literature on this issue here. I set aside skepticism over Bowles and Gintis' claim going forward as my impression is that they are on relatively firm ground.

> Because the truth-telling that is necessary to convert private to public information cannot be expected in the absence of social preferences and because public information is essential to the empirical plausibility of both the simple reciprocal altruism model and its indirect reciprocity variant, these models do not provide adequate explanations of cooperation among amoral and self-regarding individuals. (p. 70)

To me, this passage lays bare the ultimate reason why Bowles and Gintis think that explanations for cooperation based upon either reciprocal altruism or indirect reciprocity are bound to be inadequate. The reason is *not* that the free-rider identification problem is insoluble in even modestly sized social groups. Rather, it's just that solving this problem requires that group members be disposed to share (truthful) information they possess about one another's behavior.[4] I will call information about others' behavior (including others' dispositions to behave in various ways) "social information" in what follows.[5]

Consider reciprocal altruism first. As we saw above, Bowles and Gintis draw attention to the fact that, when a cooperative interaction involves more than two interactants, there is no guarantee that a given interactant will know who has cooperated and who has defected. This is because the payoff they receive does not reveal such information. The importance of this point, I suggest, isn't that it will therefore be hard or costly for group members to come to know who has cooperated and who has defected; instead, it's that often one will have no choice but to rely on the reports of others to ascertain such information (for example, those who were in a position to directly perceive the act(s) of defection). At this stage of the dialectic, Bowles and Gintis will then ask (quite rightly in my view) why it is legitimate for individual-selection accounts of cooperation to simply assume that individuals are disposed to share such information. As the last passage makes clear, Bowles and Gintis' view is that individual-selection accounts have no explanation for why individuals should be disposed to behave in this way. Proponents of group selection, by contrast, will say that this disposition reflects selection at the level of groups (i.e., groups in which individuals were disposed to share social information with one another out-survived/competed those in which individuals were not disposed to behave in such a manner).[6]

Matters are even clearer in the case of indirect reciprocity. For as Bowles and Gintis point out, reputational effects will only work to promote cooperation if group members in fact possess copious amounts of true social information about one another. Surely, some of this information can be gathered on the basis of one's own perceptual experience (including one's experiences of observing others interact), but not all of it. This is especially clear in cases where the interaction takes place in private (e.g., two individuals go off hunting together). The rest of this social information, which is a sizable package, to be sure, must be supplied by others. But again, Bowles and Gintis will want to know why we should expect group members to share social information with one another. Until individual-selection proponents have offered an explanation for how this disposition can evolve via purely individual-selection means, they will not have offered an adequate account of human cooperation.

In sum: Bowles and Gintis take the free-rider identification problem to pose a serious stumbling block for individual-selection accounts of human cooperation. In this section, I have suggested that the real crux of their argument comes down to the following: the problem of free-rider identification can only be solved provided that group members are disposed to share social information. Bowles and Gintis' view is that group-selection accounts have a

[4] In what follows, I drop the qualifier "truthful." The reader should assume that when I talk of sharing information, I mean sharing truthful or accurate information, unless I explicitly say otherwise.

[5] I use the term "social information" rather than "gossip" so as to avoid the negative connotations associated with the latter.

[6] Ultimately, one will want to see a more detailed group-selection explanation for the evolution of the disposition to share social information. The motivation for such a desire is that there would seem to be another "subversion from within" problem lurking in the background. If this disposition is, on the whole, individually costly, then those who lack the disposition will have a higher fitness than those who possess it when each belongs to the same group. What, then, is to prevent individuals who lack this disposition from overtaking the group? (I thank an anonymous referee for drawing my attention to this issue.)

I think there are a couple of things proponents of group selection can say at this point, though I do not deny that matters are complicated. Here, for example, is a sketch of one line of response which I find to be prima facie plausible. On the reasonable assumption that the costs of possessing this disposition (if indeed there are some) aren't too significant, then the difference in fitness between an individual who lacks this disposition and one who possesses it cannot be too significant either. To provide a point of contrast: this fitness difference is bound to be much smaller than that between an individual who, say, lacks the disposition to share food (and yet is not punished/ostracized by group members) and one who is disposed to share food. Hence, in the case of the disposition to share social information, it may not be that difficult for selection operating at the level of the group to offset the process of subversion from within, despite the slower pace of the former process.

On the other hand, one might think that the only fair-minded thing to say in this context is that the evolution of this disposition is actually a theoretical difficulty for *both* individual-selection and group-selection accounts (though for different reasons). (This is the view suggested to me by the above-mentioned referee.) On this view, Bowles and Gintis' argument would still be seen as raising an important evolutionary question (i.e., how could this disposition evolve?), just not one which group-selection accounts have an obviously easier time answering. I'm inclined to think that this view is a bit too strong, but I will not argue for that claim here.

ready explanation for this disposition, one which human beings clearly do possess,[7] whereas individual-selection accounts do not.

Sterelny's Criticism of Bowles and Gintis

Sterelny thinks that the free-rider identification problem has been much exaggerated by researchers like Bowles and Gintis. For Sterelny, the only real challenge for individual-selection accounts of human cooperation is why our Pleistocene ancestors would have been disposed to punish free riders (thereby incurring at least a short term cost) once free riders were identified. Because Sterelny thinks that a disposition to punish can be explained wholly within an individual-selection framework, he is unconvinced that we need to appeal to group-selection to explain how human cooperation evolved.[8]

Here is the key passage from Sterelny on Bowles and Gintis' argument:

> … Bowles and Gintis think that the problem of scale is intractable for individual-selection accounts of cooperation. … [C]ooperation is unstable even in modestly sized groups, unless free riders can be specifically targeted. That presupposes that such free riders can be identified and, once identified, punished. I do not see identification as a serious problem: I have already argued that in small, repeatedly interacting band-sized groups, agents will become well-informed about one another just by direct observation and memory. This effect is magnified by gossip, once such communicative capacities evolve. Of course, gossip is not perfectly honest and reliable, but as I have argued … multisender, multireceiver networks are well insulated against deception, and participating in those networks brings benefits of reciprocation and information pooling. No social environment is perfectly transparent, but in the social environments in which strong-reciprocity psychologies evolved in, agents were awash in information about their peers. (pp. 182–183)[9]

Sterelny puts several important ideas on the table in this passage. First, Sterelny claims that direct observation and memory go considerably further towards identifying free riders than researchers (in particular, modelers) such as Bowles and Gintis generally realize. Second, he claims that the "shape" of human communication networks can serve to significantly damp down deceptive exploitation. Third, he claims that participating in such networks produces real benefits for the individual. Let us consider each of these ideas, respectively.

I agree with Sterelny that modelers of human cooperation often make entirely unrealistic assumptions about the cognitive abilities of members of real social groups. Bowles and Gintis are no exception. A good example of this tendency, one Sterelny actually singles out, can be found in the first passage from Bowles and Gintis cited above. Bowles and Gintis tell us that onlookers "may think a retaliatory defection is a free-riding defection, inviting further retaliatory defections" (2011, p. 64). Regarding this assumption, Sterelny writes that:

> As the models represent [cooperative] interactions, onlookers see no intrinsic difference between one agent defecting against a second, and that agent punishing the second. In real social environments, these are very different. For example, communication among those interacting will be very different. The defector will be attempting to persuade his victim that he is in fact cooperating; in the case of punishment, the agent will be denouncing his target. Denunciation will not tell us whether the punishment is just. But onlookers will nonetheless see two very different interactions. (p. 184)

I am inclined to side with Sterelny here. On the other hand, his concession that observation alone will not suffice to tell us whether the one individual is justified in punishing the other is an important one. (If an onlooker is unable to tell whether the punishment is just or unjust, he or she will not know which attitude it is appropriate to adopt towards the interactant who is doing the punishing.) More generally, though, I contend that even if Bowles and Gintis do underplay the role of direct observation and memory in identifying free riders, the core of their argument remains intact. If the free-rider identification problem is to be solved in all but the most intimate social settings, it is crucial that individuals are disposed to share social information, and to abstain from deceptive manipulation.

[7] See, for example, Emler (1992, 1994), Dunbar et al. (1997), and Wiessner (2005). Emler sampled from a wide range of conversational contexts and found that about 70 % of time spent conversing involves the transmission of social information. Following this study up, Dunbar et al. examined the content of conversations taking place between students in a university refectory. They found that about 60 % of conversational time is spent on social information. Finally, based on 308 multi-hour recordings of conversations between Ju/'hoansi!Kung individuals, Wiesnner reports that 56 % of conversations included criticism of group members, and (only!) 7 % praise of group members.

[8] It should be noted that Sterelny is not in general opposed to the idea that group selection has played a role in human evolution (see 2012, p. 178, for example); he simply rejects the idea that we need to appeal to group selection to explain how the disposition to cooperate evolved in humans.

[9] All page numbers following block quotes in this section refer to Sterelny (2012).

Surely, this will at least be true when many cooperative interactions only involve a subset of group members, and take place in private. My guess is that Sterelny would concur, as he himself draws attention to the role played by gossip in solving the free-rider identification problem.

Sterelny's claim that the shape of a communication network can serve as an honesty-enforcing mechanism is, in my view, spot on.[10] In particular, Sterelny points out that when the receiver of a message can check the message's truth by consulting several or more group members, it will often be in the sender's interest to communicate truthfully, if he communicates at all. Deception in such a context is not only unlikely to be effective; it can indeed serve to give one a bad reputation. In addition, Sterelny points out that as the number of receivers that a sender communicates with goes up, so too does the chance that his deception will be detected.[11] This is because it becomes less likely that all of the receivers involved will be ignorant in a way that makes them vulnerable to the deception. For example, if a sender falsely relates that a food patch has been exhausted, he may well persuade those individuals who have not visited the patch of late. But he will *not* succeed in persuading those who have just visited the patch themselves. So, if the sender's deception is broadcast widely, there is a good chance that it will be received not only by individuals of the former sort, but also by those of the latter. Together, Sterelny claims that these two features of human communication networks will work to undercut the temptation of self-regarding individuals to deceive their group mates.

Let us assume that ancestral human groups were in fact characterized by the sort of communication networks Sterelny has in mind here (i.e., many-to-many networks). If so, then Sterelny does have an explanation for why the individuals belonging to such groups would have been disposed to abstain from the spreading of false information, an explanation that is at the very least prima facie plausible. This is key, as widespread deception concerning others' behavior certainly seems like it would render reciprocal altruism and indirect reciprocity incapable of supporting cooperation. But, one might wonder, is the many-to-many nature of such communication networks really something which individual-selection accounts of cooperation are entitled to help themselves to? I take it that part of what makes it the case that a given communication network is many-to-many is that the nodes belonging to the network are disposed to share information with one another. Would such a disposition have served to confer a net expected benefit upon the individual? And more to the present point, would a disposition for sharing *social information* in particular have served to confer a net expected benefit on the individual? Bowles and Gintis would claim that a disposition to share social information would *not*, on the whole, have been individually advantageous. Sterelny thinks otherwise. He writes:

> It is true that sharing information via gossip is indeed an *N*-player cooperation problem. But the costs and benefits of informational cooperation do not mirror those of ecological cooperation. Costs are lower; benefits are higher. The benefits of informational sharing increase as *N* increases, for Condorcet-like effects increase the reliability of consensus. And the more gossip is multisourced and multitargeted, the less likely it is to be deceptive manipulation. Moreover, in contrast to sharing material resources, the costs of sharing information do not increase with the number of agents aided. But the number of reciprocation sources does rise with informational targets. You can help more people—and hence legitimately expect help back from more people—for a fixed cost. Finally, Nowak and Sigmund (2005) point out that experimental evidence shows that agents are sensitive to others' reputation and expect others to be sensitive to their own reputation. (p. 183)

In my view, this passage cuts to the very heart of the debate at hand. Sterelny suggests that sharing social information is correctly viewed as posing a kind of cooperation problem in itself. On this point, Bowles and Gintis would surely agree. However, Sterelny *also* thinks that individual-selection accounts have a ready explanation for why individuals would be disposed to share social information. I think matters are a good deal more complicated than Sterelny realizes.

First, Sterelny again draws attention in this passage to how the many-to-many shape of communication networks can serve to damp down deception. I am happy to go along with the idea that the communication networks via which social information flowed in ancestral human groups were in fact of this sort. However, the question we are presently interested in is *why* the individuals belonging to these groups

[10] Sterelny does not claim that this is the only honesty-enforcing mechanism, of course. In particular, he approvingly cites the work of Sperber (e.g., Sperber 2000, 2001) on the role of folk epistemology in helping us to evaluate the reliability of various sources. What Sterelny disagrees with Sperber (and many others) on is that epistemic vigil against deception is the primary means of keeping deception in check. Sterelny's point is that the shape of human communication networks already does much of this work for us.

[11] The communal meetings which many hunter-gathers hold at the end of each day provide us with a vivid illustration of a communicative situation involving very many receivers. For example, Hames and Vickers (1982) report that, "When [Siona-Secoy hunters, an indigenous Amazonian people] return from the forest, they gather to discuss their various successes and failures, the signs of game encountered, the location of fruiting plants favored by particular species, and the difficulties of tracking, stalking, and pursuit" (p. 368). Other examples can be found in Ray (1963), Marshall (1976), and Tanner (1978).

were disposed to share social information in the first place. Why were they willing to share social information as opposed to simply keeping information of this sort to themselves? Appealing to network shape in order to explain the existence of such a disposition not only seems to get the order of things backwards (as I just mentioned, a network for transmitting social information *counts* as many-to-many in virtue of the fact that the network's nodes are disposed to share social information); it also strikes me as a bit of a red herring. The *truthfulness* of the information that in fact flows through some communication network is one thing; why there is a network there *at all* is another. (Of course, the two are not entirely unrelated. In particular, should deception become rampant, receivers will presumably stop listening.)

Second, Sterelny says that Condorcet-like effects help to explain why individuals would have been disposed to share information. Here he is alluding to the Condorcet Jury Theorem which tells us that if each "juror" has a more than 50 % chance of being correct, then the probability of the majority view being correct goes up with the number of jurors. This suggestion is on the right track in the sense that it is the right kind of reason for explaining why individuals would have been disposed to share information. Sterelny's idea is that it would have been in individuals' own self-interest to pool their information so as to produce Condorcet-like effects. In this way, each individual would have regularly gained access to more accurate judgments, which is something that surely would pay at the individual level. I agree with Sterelny that the logic behind the Condorcet theorem provides a compelling explanation for why ancestral humans would have been disposed to share *some* kinds of information; what I am skeptical of is that this logic applies in the case of sharing social information. If, for example, a group must decide whether they have a better chance of surviving if they move to a new location rather than staying put (suppose they are faced with a severe drought), then the Condorcet theorem tells us that it is in the self-interest of each individual to speak up. What is far from clear to me is that this sort of case is a good general model for thinking about the sharing of social information. To give a concrete example: suppose that A discovers that B is hoarding food. Can a disposition for A to inform members of her group about anti-social behavior of this kind really be explained, even in some ultimate sense, by A's gaining access to a more accurate judgment about B's personality type? At the very least, we are surely owed another layer of detail here, one focused specifically on the sharing of social information, not information in general.

Sterelny offers two further considerations in support of his claim that a disposition to share social information can be explained via individual selection alone. The first of these concerns *reciprocal* information sharing. Put roughly, the idea here is that if A shares information with B at time t, then B is more likely to share information with A at a future time t' than B would have been otherwise. The real importance of this point for Sterelny in the present context is that (according to him) the expected costs of sharing some piece of information remain fixed as the number of group members with whom this information is shared increases. Consequently, it will be quite easy for the expected benefits of sharing this information to outweigh the expected costs. I will look at this line of reasoning in more detail in the next section.

The other consideration Sterelny raises here concerns individuals' reputations. Sterelny cites the experimental work of Nowak and Sigmund (2005) showing that individuals condition their behavior (cooperate versus defect) on the reputations of those with whom they interact, and expect others to do the same. There is a narrow and broad way of thinking about results like Nowak and Sigmund's. Looked at narrowly, what these results tell us is that when group members have access to accurate information about how one has behaved in past cooperative settings, it will often be in one's self-interest to behave cooperatively, even when the chance of a repeat interaction is low (or possibly even nil). Looked at broadly, these results simply drive home something which is clear to each of us from everyday life, namely, humans typically reflect upon what they know about individuals with whom they are interacting, or with whom they are contemplating interacting, in deciding on a specific course of action. On this view, the importance of reputation extends far beyond prisoner's dilemma-style interactions; in particular, reputation is a key factor influencing *partner choice*.[12] In choosing a partner, we not only attend to how some candidate partner has behaved towards ourselves and others in the past; we may also factor in general personality traits (e.g., is brave, is hot-tempered, etc.) as well as expertise (e.g., is a plant expert, is mechanically inclined, etc).

It is very probable that reputational effects (construed broadly) will figure centrally into any plausible individual-selection account of informational sharing.[13] In sharing information with one's group members, one can often indicate to others that one possesses certain otherwise hidden attributes that make one a desirable partner for future activities, activities that will produce real benefits for the individuals involved.

[12] For state-of-the-art work on partner choice in the human domain, see Baumard et al. (2013).

[13] I am indebted to an anonymous referee for this insight. Without his/her detailed comments, I would not have appreciated the importance of partner choice in this context. The material presented in the remainder of this section draws heavily upon those comments. The main way in which my discussion goes beyond that suggested to me by the referee is in showing how the precise *content* of the information that is shared can signal the possession of specific personality traits (e.g., is very knowledgeable about animal behavior).

To see how this might go, consider the following case. Consider an individual A who is disposed to share information with his group mates that he has gathered while out hunting alone. On this particular day, A observes tracks and other signs left by a prized game animal. On the basis of these signs, A is able to work out the species, size, health, and likely whereabouts of the animal. Upon returning to camp, A, as usual, shares this newly acquired information with his fellow hunters.

From an individual-selection perspective, A's disposition can indeed appear quite puzzling. The reason being, such acts of informational sharing on A's part would seem to produce real costs for A.[14] For example, when A informs others that he has observed signs of a high-quality game animal, and furthermore tells them where he believes the animal to be located, A lowers the chance that he will be the first to lay eyes upon the animal and make the kill. This is because other hunters now strike off in pursuit of the prey.[15]

Wouldn't A be better off *not* having such a disposition? It is surely tempting to think so. This impression is made all the stronger, I think, by the fact that were A to frequently withhold certain pieces of valuable information from the group instead, others would generally not be in a position to directly perceive such acts of defection on A's part. In this regard, refusing to share information is importantly different from refusing to share material resources like food or to provide physical assistance to others (when one is visibly strong).

The problem with this way of thinking is that it does not take into account the benefits our individual A would be likely to see as a result of having such a disposition. In regularly sharing valuable information that he has gathered while out hunting, A not only builds and maintains a reputation as a helpful and generous member of the group; in addition, *he cultivates the perception in others that he is keen-minded and a strong natural historian* (not just anyone can recognize animal tracks, infer the quality of the animal, infer its direction of movement, etc).[16] Of course, these are just the sorts of (otherwise hidden) attributes that make one an excellent partner for a wide range of hunter-gather activities. Thus, while A's disposition may well have real individual costs, these costs (one might think) are ultimately outweighed by the reputational gains that flow from A's informational sharing. A is more likely to be chosen for future profitable activities by members of his group than he would have been had he not shared information in this way.

More than a few questions arise at this point regarding the details of this sort of explanation for informational sharing. Unfortunately, it is beyond the scope of this article to go down that path. What I plan to do instead is consider the applicability of this sort of explanation to the case of sharing social information.

Sharing Social Information

In this section, I turn to the expected costs and benefits of having a disposition to share social information.

It is helpful to begin by stating what is common ground between Bowles and Gintis, and Sterelny. First, both agree that humans are disposed to report on the behavior of others. (It can hardly be doubted that humans relish the opportunity to engage in some "good" gossip). Second, both (presumably) agree that there is a widely obeyed norm of truth-telling as concerns the transmission of social information. Finally, both would agree that these two facts form (at least part of) the foundation for an explanation of how the free-rider identification problem is solved in all but the most intimate social settings. Were humans *not* so disposed, then free riders would too often go unidentified and so unpunished. Consequently, cooperation within the group would begin to break down and eventually disappear altogether.

On what do the two parties *disagree*, then? As I see matters, what they ultimately disagree on is whether an individual-selection explanation for these facts about human communication can be provided. Bowles and Gintis think that no such explanation is forthcoming. Sterelny, by contrast, thinks that proponents of individual-selection have a ready explanation.

Below, I set aside issues about truthfulness. I agree with Sterelny that the opportunities individuals encounter to engage in successful deception have been much overestimated by researchers. It *is* true, however, that those features of human communication networks which serve to damp down deception (the fact that information is multi-targeted and multisourced, as Sterelny says) should not be taken for granted by proponents of individual-selection, but I will not consider this issue further here. Instead, in the remainder of this article, my focus will simply be on why individuals are disposed to share social information at all.

[14] To be clear: I here assume that the hunter is perfectly capable of bagging the prey without enlisting the help of others. If he weren't, then his sharing of such information, at least with some of his group members, wouldn't be at all surprising. I thank an anonymous referee for this clarification.

[15] Of course, even if meat is typically shared amongst the group, it is not hard for us to imagine reasons why each hunter prefers to be the one who actually bags the prey. For example, he who bags the prey may enjoy an uncontested claim to some especially highly valued portion of the meat.

[16] Marlowe (2010), for example, relates how Hadza people attempt to signal that they have "good eyesight, coordination, strength, knowledge, endurance, [and] bravery" (p. 231). Marlowe's discussion concerns the sharing of particular foods (it may take great bravery, say, to secure a certain food type), but shared information (if true) could serve essentially the same purpose.

In determining whether an individual-selection explanation can ultimately be provided for this disposition, we need to inquire into the expected costs and expected benefits of having such a disposition. If it is plausible that the expected benefits would have outweighed the expected costs, given the kind of social groups our Pleistocene ancestors lived in, then I take it that a disposition to share social information can indeed be explained along individual-selection lines. But if it is plausible that the expected costs would have *failed* to outweigh the expected benefits, then I take it that some other kind of explanation must be provided (i.e., either a group-selection explanation or else some kind of nonadaptive explanation[17]).

Let us consider the expected costs first. To be clear, what we are looking for are expected costs that *typically* accompany acts of sharing social information, not ones that arise from the details of highly specific social interactions. (For example, we *don't* want to factor into our analysis at this stage cases in which one would be violating the expectations of one's communicative partners, were one to share some piece of social information with them.) I think there is at least one major source of expected costs that individuals incur by sharing social information, though these expected costs are limited only to the sharing of *negative* social information, that is, information about socially disapproved of behaviors. (I think negative information is pretty clearly the more important of the two types of social information—the other being *positive* social information, that is, information about socially approved of behaviors—when it comes to sustaining cooperation, though I won't argue for that claim here.) The expected costs stem from the relation that the sharer bears to the *target* of the information being shared, whereby "target" I mean the individual whose behavior the information is about.

Suppose A shares some negative and therefore potentially damaging information about individual B with group members C, D, E, and so on. In so doing, A incurs a risk. The risk is this: should B come to know that A has shared this information with members of the group, B may well retaliate against A. Of course, one way B may learn of this fact is that one of the individuals A informs (or one of the individuals that is informed by an individual that A informs, etc.) simply tells B. That is, some individual may alert B (perhaps in an attempt to win favor with B) that A is sharing information about B's antisocial behavior. Alternatively, if the information in question is something *only* A would know, as would be the case when the information concerns B's behavior in the context of some private interaction between A and B, then the fact that this information has spread beyond A will, by that very fact, alert B that A (and possibly others) have shared this information. Now, B's retaliation against A may assume a variety of forms. One thing B might do is simply break off a preexisting social or utilitarian arrangement with A, an arrangement that has brought A real benefits in the past. (For example, perhaps B has been a useful hunting partner to A in the past, and now B refuses to hunt with A; or perhaps B is an expert weaponry maker, and now B will no longer trade with A.) However, B might not be so coolheaded. In the extreme case, B's retaliation against A may take the form of full-fledged physical violence against A. Whether retaliation of the latter sort would ultimately be more costly for A than retaliation of the former will of course depend on a range of variables. At any rate, the important thing to note is just that, in sharing negative social information with one's group mates, an individual thereby incurs the possibility of real target-induced costs. Moreover, it would appear that the expected costs will in general *rise* with the number of individuals that one informs, assuming that sharing this information with a larger number of individuals increases the chance of the target learning that one has shared negative information about him.

What about the expected benefits that typically accompany the sharing of social information? In the last section, we mentioned two possible sources of expected benefit, namely, reciprocal informational sharing and reputational gains. Let us take these, respectively. Suppose that A shares information with C about B's (let's say, antisocial) behavior. In so doing, A may be contributing a real benefit to C; C may have been considering B as a partner for a future activity, one requiring trust and honesty, and now C knows she should avoid B. C recognizes this good deed on A's part and is inclined to pay it back. More specifically, C now provides A with some valuable social information (or perhaps some other kind of information[18]) at a future point

[17] The fondness which we show for sharing social information certainly stands out in comparison with our feelings towards sharing other kinds of information. One might think that this constitutes strong evidence that our disposition to share such information has been shaped by one or another selective force. Put differently, one might think that it is very unlikely that we would show such a fondness for sharing social information by chance alone or as a by-product of some other psychological trait. If this is correct, then we should obviously be quite skeptical of any nonadaptive explanation for our disposition to share social information. This is really a topic deserving of an in-depth treatment of its own, however. I take no definite stance on the relative plausibility of adaptive versus nonadaptive explanations for this disposition in the present work.

[18] The possibility that acts of sharing social information might be "paid back" by acts of sharing other kinds of information (e.g., ecological information), or just information more generally, is an important one. There is not room to consider this possibility here, however, and so I set it aside in what follows. What I want to focus on is the possibility that, in sharing social information, an individual gains access to a specific domain of information (social information) that she would lack access to if she did not herself share social information.

in time. The contrast case to consider here is one in which A instead withholds this information from C, and C, as a result, withholds valuable social information from A in the future. Assuming this kind of pattern generalizes, it is easy to see that A gains real benefits from a disposition to share social information. A can reliably expect to receive social information back from all, or at least most, of those individuals whom she has aided with social information. Moreover, at least in the case of positive social information, we would intuitively expect (apparently incorrectly, as it turns out—see below) the expected costs of sharing such information to remain fixed as the number of individuals one informs goes up.

Moving on to reputational gains: it has been suggested that, by sharing social information, one can send a signal to others that one enjoys a place of social power or influence within one's group (e.g., Kurland and Pelled 2000; Dunbar 2004; Foster and Rosnow 2006; Guerin and Miyazaki 2006). The idea here is that only those individuals who enjoy a place of social power within the group will have much social information to share; hence, one should indeed share this sort of information, and share it widely, if one has it. This explanation makes use of the same kind of logic that we used to explain a disposition to share ecological information (or more precisely, information acquired while out hunting) in our example from the last section. That is, in both cases, the acts of informational sharing which flow from the disposition work to cultivate the perception in others that the sharer has certain attributes that make him or her an attractive partner for some range of activities (in the case at hand, the revealed attribute is a high degree of social power).

I claim that we cannot hope to settle the issues at play by simple reflection on the possible costs and benefits of sharing social information in the sorts of groups our Pleistocene ancestors lived in. We can imagine reasons why having such a disposition might work to the individual's benefit, but we can also imagine reasons why it could hurt the individual. Unfortunately, what little experimental work exists in this area hardly gives us a clear verdict. At most, what this work entitles us to conclude is that being disposed to share social information at high frequencies (to report on others' behavior often, and to a wide audience) is likely detrimental to the individual on the whole (Turner et al. 2003; Farley 2011). (Surprisingly, this appears to be true even when much of the social information one shares is of the *positive* variety.) In a word, those who share social information at high frequencies are apparently perceived as being less trustworthy, less likeable, and less socially powerful than those who share social information at low frequencies. Of course, it is not hard to see how such perceptions on the part of others might translate into real costs for the individual (and this is to say nothing of the target-induced costs mentioned above). Most obviously: having a reputation for being a "big mouth" may well deter others from sharing social information with you, and may even cause you to be excluded altogether from social interactions where sensitive information is likely to be divulged.[19] What this work does *not* tell us is how a disposition to share social information at a more moderate level might bear upon individual fitness (a point Farley herself makes in discussing the shortcomings of her study), although it does seem to rule out simpleminded versions of the idea that sharing social information can be explained by status enhancement.

More theoretical and (in particular) empirical work[20] is definitely needed in this area before we can hope to definitively answer whether the human disposition to share social information can be explained from a purely individual-selection perspective. To begin with, we will presumably need a much more precise description of the patterns of informational sharing which individuals actually engage in. (Since there are bound to be individual as well as cultural differences as regards such sharing, we will eventually want to replace talk of "the disposition to share social information" with more detailed language. I bracket this complication below, as I have done so far in this article.) More specifically, we will need to know, among other things:

- How is individual decision making about whether to share social information sensitive to the relationship between a potential *sharer* and the *target* of that information?
- How is such decision making sensitive to the relationship between a potential sharer and his or her *communicative partner(s)*?[21] and
- How is such decision-making sensitive to the *specific content*[22] of information that is a candidate for sharing?

[19] The reason I did not include these expected costs earlier in the discussion in this section is that they flow from a rather extreme form of the disposition to share social information.

[20] I construe "empirical" broadly, so as to include, for example, looking to preexisting ethnographic accounts to gain a better understanding of how social information is exchanged in small-scale societies.

[21] The experimental work of Turner et al. cited in the foregoing paragraph did look at how perceptions of the sharer were affected by relation to the sharer (i.e., whether the sharer was a friend or a stranger). Surprisingly, they found that sharing social information, regardless of whether it was positive or negative, tended to adversely affect perceptions of the sharer for both types of relationships. What I am asking about here is different, though. I am asking how individuals factor in their relationship to their communicative partner(s) in deciding what social information, if any, to share with the latter. A more general question here is how such decision making is sensitive to the personality traits of one's communicative partner(s).

[22] I have more in mind than simply whether the information is positive versus negative.

In short, we will need a fuller picture of the factors that enter into individuals' decisions about which social information to share and with whom to share it. But in addition to these questions focused on the sharer side of things, we will also need to know how acts of sharing tend to affect the behavior of those with whom social information is shared. In particular:

- If A shares social information with B, is it *actually* the case that B is more likely to share social information with A in the future than if A hadn't shared such information with B?
- If A regularly (though still selectively in some sense) shares social information with his or her group mates, is A *actually* more likely to be chosen as a partner for future activities that pay off for A?

With even partial answers to these questions, we will be in a much better position to evaluate the claim that the disposition to share social information can be explained via individual selection alone. That is, we will be better able to assess the claim that, for an individual belonging to a social group of the sort our Pleistocene ancestors lived in, the expected benefits to the individual of having such a disposition would have outweighed the expected costs.

Where have we gotten to, then? We began by clarifying Bowles and Gintis' argument from the free-rider identification problem to a group-selection account of the evolution of human cooperation. I claimed that the real crux of their argument is not that this problem is exceedingly difficult to solve in even modestly sized social groups, but rather that its solution in such groups requires that individuals are disposed to share social information with one another. This is the explanatory challenge that I think Sterelny underestimates. While Sterelny *does* realize that proponents of individual-selection accounts cannot simply take such a disposition for granted, the brief explanation he gives leaves too many important questions unanswered. Crucially, however, this is *not* to say that we should be convinced by Bowles and Gintis' argument; they are unduly skeptical, I think, of the possibility that an adequate individual-selection explanation for the disposition to share social information can be provided. I leave the examination of this possibility for another day.

Acknowledgments I would like to thank the two anonymous referees for this journal whose excellent feedback greatly improved this article.

References

Alexander R (1987) Biology and human affairs. University of Washington Press, Seattle

Axelrod R, Hamilton W (1981) The evolution of cooperation. Science 211:1390–1396

Baumard N, André J, Sperber D (2013) A mutualistic approach to morality: the evolution of fairness by partner choice. Behav Brain Sci 36:59–78

Bowles S, Gintis H (2011) A cooperative species: human reciprocity and its evolution. Princeton University Press, Princeton

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dunbar R (2004) Gossip in evolutionary perspective. Rev Gen Psychol 8:100–110

Dunbar R, Marriott A, Duncan N (1997) Human conversational behavior. Hum Nat Int Bios 8:231–246

Emler N (1992) The truth about gossip. Soc Psychol Sect Newsl 27:23–37

Emler N (1994) Gossip, reputation and social adaptation. In: Goodman R, Ben-Ze'ev A (eds) Good gossip. University of Kansas Press, Lawrence, pp 119–140

Farley S (2011) Is gossip power? The inverse relationships between gossip, power, and likeability. Eur J Soc Psychol 41:574–579

Foster E, Rosnow R (2006) Gossip and network relationships. In: Kirkpatrick D, Duck S, Foley M (eds) Relating difficulty: the processes of constructing and managing difficult interaction. Erlbaum, Mahway

Guerin B, Miyazaki Y (2006) Analyzing rumors, gossip, and urban legends through their conversational properties. Pscyhol Rec 56:23–34

Hames R, Vickers W (1982) Optimal diet breadth theory as a model to explain variability in Amazonian hunting. Am Ethnol 9:358–378

Kurland N, Pelled L (2000) Passing the word: toward a model of gossip and power in the workplace. Acad Manag Rev 25:428–438

Marlowe F (2010) The Hadza hunter-gatherers of Tanzania. University of California Press, Berkeley

Marshall L (1976) The !Kung of Nyae Nyae. Harvard University Press, Cambridge

Nowak M, Sigmund K (1998) Evolution of indirect reciprocity by image scoring. Nature 393:573–577

Nowak M, Sigmund K (2005) Evolution of indirect reciprocity. Nature 437:1291–1298

Ray V (1963) Primitive pragmatists: the Modoc Indians of Northern California. University of Washington Press, Seattle

Sperber D (2000) Metarepresentation in an evolutionary perspective. In: Sperber D (ed) Metarepresentations: a multidisciplinary perspective. Oxford University Press, Oxford, pp 117–137

Sperber D (2001) An evolutionary perspective on testimony and argumentation. Philos Top 29:401–413

Sterelny K (2012) The evolved apprentice: how evolution made humans unique. MIT Press, Cambridge

Sugden R (1986) The economics of rights, co-operation and welfare. Basil Blackwell, Oxford

Tanner A (1978) Divinations and decisions: multiple explanations for Algonkion scapulimancy. In: Schwimmer E (ed) The yearbook of symbolic anthropology. McGill-Queens University Press, Montreal, pp 89–101

Trivers R (1971) The evolution of reciprocal altruism. Q Rev Biol 46:35–57

Turner M, Mazur M, Wendel N, Winslow R (2003) Relational ruin or social glue? The joint effect of relationship type and gossip valence on liking, trust, and expertise. Commun Monogr 70:129–141

Wiessner P (2005) Norm enforcement among the Ju/'hoansi Bushmen: a case of strong reciprocity? Hum Nat Int Bios 16:115–145